SIAM J. NUMER. ANAL.
Vol. 29, No. 5, pp. 1244–1260, October 1992

NUMERICAL WAVE PROPAGATION IN AN ADVECTION EQUATION WITH A NONLINEAR SOURCE TERM*

D. F. GRIFFITHS†‡, A. M. STUART‡§, AND H. C. YEE‡

Abstract. The Cauchy and initial boundary value problems are studied for a linear advection equation with a nonlinear source term. The source term is chosen to have two equilibrium states, one unstable and the other stable as solutions of the underlying characteristic equation. The true solutions exhibit travelling waves which propagate from one equilibrium to another. The speed of propagation is dependent on the rate of decay of the initial data at infinity. A class of monotone explicit finite-difference schemes are proposed and analysed; the schemes are upwind in space for the advection term with some freedom of choice for the evaluation of the nonlinear source term. Convergence of the schemes is demonstrated and the existence of numerical waves, mimicking the travelling waves in the underlying equation, is proved. The convergence of the numerical wave-speeds to the true wave-speeds is also established. The behaviour of the scheme is studied when the monotonicity criteria are violated due to stiff source terms, and oscillations and divergence are shown to occur. The behaviour is contrasted with a split-step scheme where the solution remains monotone and bounded but where incorrect speeds of propagation are observed as the stiffness of the problem increases.

Key words. numerical wave propagation, heteroclinic orbits

AMS(MOS) subject classifications. 65M06, 35L60

1. Introduction.

Consider the following initial boundary value problem for the scalar $u(x,t)$:

$$u_t + u_x = f(u), \qquad x \in \mathcal{R}, \tag{1.1}$$

$$u(x,0) = a(x), \qquad x \in \mathcal{R}. \tag{1.2}$$

The function $f(u)$ is assumed to satisfy the following three conditions:

(I) $f(u) \in C^2([0,1],\mathcal{R})$;
(II) $f(0) = f(1) = 0$ and $f(u) > 0$ for $u \in (0,1)$;
(III) $f'(0) > 0, f'(1) < 0$.

It is assumed that the initial data satisfies

(IV) $a(x) \in [0,1]$, $x \in \mathcal{R}$.

In some situations the case $a(x) \equiv 1$ for $x \leq 0$ is considered and the problem is then equivalent to an initial boundary value problem posed on $x,t \geq 0$ with $u(0,t) = 1$. We refer to this initial boundary value problem as (IBVP) and to the general Cauchy problem as (C). A typical example of a nonlinear source term satisfying (I)–(III) is $f(u) = u - u^2$.

We consider the following explicit finite-difference approximation for (1.1), (1.2):

$$U_j^{n+1} = cU_{j-1}^n + (1-c)U_j^n + \Delta t g(U_{j-1}^n, U_j^n), \tag{1.3}$$

$$U_j^0 = a(j\Delta x). \tag{1.4}$$

*Received by the editors April 9, 1991; accepted for publication October 3, 1991.
†Department of Mathematics and Computer Science, The University, Dundee DD1 4HN, United Kingdom.
‡NASA Ames Research Center, Moffett Field, California 94035.
§School of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom. Current address: Division of Applied Mechanics, Stanford University, Stanford, California 94305.

Here $U_j^n \approx u(j\Delta x, n\Delta t)$ and $c = \Delta t/\Delta x$. We now detail our assumptions about the scheme (1.3), using the notation

$$h(u,v) = cu + (1-c)v + \Delta t g(u,v).$$

(i) $g(U,U) = f(U)$;
(ii) $g \in C^2(\mathcal{R}^2, \mathcal{R})$; $g_u(U,U) + g_v(U,U) = f'(U)$;
(iii) $h_u(u,v) \geq 0 \; \forall \, u, v \in [0,1]$;
(iv) $h_v(u,v) \geq 0 \; \forall \, u, v \in [0,1]$.

Assumptions (i) and (ii) are consistency assumptions whereas assumptions (iii) and (iv) are constraints on the parameters of the scheme. In the case where the source term $f(u) \equiv 0$, conditions (iii) and (iv) reduce to the standard CFL condition $c \in [0,1]$. It is worth noting that the monotonicity conditions (iii) and (iv) are reminiscent of those employed by [Harten, Hyman, and Lax, 1976] in schemes designed for scalar conservation laws. Some schemes satisfying the criteria (i)–(iv) are described in §5.

Currently there is a great deal of interest in the numerical approximation of systems of conservation laws with stiff nonlinear source terms. Such problems arise as models of nonequilibrium gas dynamics and, in particular, in the study of transatmospheric vehicles. It is our purpose to analyse such problems by considering the very simplest hyperbolic problem with nonlinear source, namely, (1.1), (1.2). Earlier work by [Colella, Majda, and Roytburd, 1986] and [LeVeque and Yee, 1990] shows that, when insufficient spatial or temporal resolution is employed, spurious phenomena can arise in the study of such problems. Specifically, for split-step schemes, they both find travelling waves with grid-sensitive propagation speeds. In contrast, we show that if sufficient resolution is employed the numerical method exhibits travelling waves that are close in form and speed to those of the underlying continuous problem. The relationship between the numerical speed of propagation and the true speed is studied in detail and convergence of the numerical wave-speed to the true wave-speed established.

The breakdown of these desirable properties is studied when (iii) and (iv) are violated, due to a stiff source term. It is shown that oscillations occur and, for increasing stiffness, divergence of the scheme. This behaviour is contrasted with a split-step scheme (similar to those studied in [Colella, Majda, and Roytburd, 1986] and [LeVeque and Yee, 1990]) where monotonicity and boundedness are preserved, but where spurious propagation speeds are observed for stiff problems.

In §2 we consider (IBVP) and analyse the convergence of the scheme. Section 3 contains analysis of (C); the existence of numerical travelling waves is proved and their behaviour studied for increasing stiffness. Section 4 is concerned with the split-step scheme and the ideas are similar to those in §3. Finally, in §5, we consider some numerical methods satisfying conditions (i)–(iv) and present results which illustrate the preceding analysis. It is our hope that the analysis contained in this paper can be extended to, or at least used to motivate analogous questions in, more complicated problems involving nonlinear convection and shock formation in systems of conservation laws.

2. Asymptotic behaviour on a fixed spatial interval and convergence. In this section we consider (IBVP) driven by the boundary condition $u(0,t) = 1$. We study the boundedness of the solution, its monotonicity, its asymptotic properties on any fixed interval and its convergence. The error estimate obtained in the convergence analysis is uniformly valid in time.

Note that, since $a(x) \in [0,1]$, explicit solution of the problem along characteristics shows that the true solution remains in the interval $[0,1]$; hence it is desirable that the numerical method shares this property.

LEMMA 2.1. *$U_j^n \in [0,1]$ for $j \geq 0$ and $n \geq 0$.*

Proof. The proof is by induction. We have $U_j^0 \in [0,1]$ for $j \geq 0$. Assume that $U_j^n \in [0,1]$ for $j \geq 0$. From (1.3) we have for $j \geq 1$, using (iii), (iv), (i), and (II):

$$U_j^{n+1} \leq c + (1-c)U_j^n + \Delta t g(1, U_j^n) \leq 1.$$

Similarly, for $j \geq 1$,

$$U_j^{n+1} \geq (1-c)U_j^n + \Delta t g(0, U_j^n) \geq 0.$$

In addition $U_0^{n+1} = 1$. Thus $U_j^{n+1} \in [0,1]$ for $j \geq 0$, and the inductive step is complete. □

A second important property is monotonicity—if the true solution is monotone initially, then it remains monotone for all time. This property is shared by the numerical method.

LEMMA 2.2. *If $U_{j-1}^0 \geq U_j^0 \; \forall j \geq 1$ then $U_{j-1}^n \geq U_j^n \; \forall j \geq 1$ and all $n \geq 0$.*

Proof. Let $\delta_j^n = U_{j-1}^n - U_j^n \; \forall j \geq 1$. Again we proceed by induction. We have that $\delta_j^0 \geq 0 \; \forall j \geq 1$. Let $\delta_j^n \geq 0 \; \forall j \geq 1$. From (1.3) we have

$$U_{j-1}^{n+1} - U_j^{n+1} = h(U_{j-2}^n, U_{j-1}^n) - h(U_{j-1}^n, U_j^n).$$

Hence, for $j \geq 2$, the mean value theorem yields

$$\delta_j^{n+1} = h_u(\eta, \xi)\delta_{j-1}^n + h_v(\eta, \xi)\delta_j^n.$$

Using Lemma 2.1 we deduce that $\xi, \eta \in [0,1]$. Hence, using (iii) and (iv), $\delta_j^{n+1} \geq 0$ for $j \geq 2$. It remains to consider $j = 1$. Now, $\delta_1^{n+1} = u_0^{n+1} - u_1^{n+1} = 1 - u_1^{n+1} \geq 0$ (using Lemma 2.1). This completes the induction. □

For (IBVP) the true solution satisfies $u(x,t) \to 1$ as $t \to \infty$ for any fixed $x \geq 0$. This property is inherited by the numerical schemes considered.

THEOREM 2.3. *Let $0 \leq j \leq J$. Then $U_j^n \to 1$ as $n \to \infty$.*

Proof. Throughout the proof $0 \leq j \leq J$. The first step is to establish that $\exists N(J)$ such that $U_j^n > 0$ for $n \geq N(J)$. From (1.3), using Lemma 2.1, we find that $U_j^{n+1} \geq cU_{j-1}^n$ and hence that, for $0 \leq k \leq J$, $U_k^{N(J)} \geq c^k U_0^{N(J)-k}$ provided that $N(J) \geq J$. Noting that $U_0^n = 1$ the result follows.

Now let

$$U_{\min}^n = \min_{0 \leq j \leq J} U_j^n.$$

Then, by monotonicity and (i),

$$U_j^{n+1} \geq U_{\min}^n + \Delta t f(U_{\min}).$$

Hence, in particular,

$$U_{\min}^{n+1} \geq U_{\min}^n + \Delta t f(U_{\min}^n).$$

For $n \geq N(J)$ we have $U_{\min}^n > 0$ and hence, for $U_{\min}^n < 1$,

$$U_{\min}^{n+1} > U_{\min}^n.$$

Thus the minima of U_j^n form an increasing sequence with 1 as the only possible limit. By Lemma 2.1 we know that $U_{\min}^n \leq 1$, and hence the result follows. □

We now prove convergence of the scheme (1.3), (1.4). The convergence result we obtain involves an error constant which is uniform in time on any compact interval in space. The method of proof is motivated by that used by [Larsson, 1989] in his analysis of finite element methods for semilinear parabolic problems; see also [Sanz-Serna and Stuart, 1990] for an application to finite-difference methods for semilinear parabolic problems. Throughout the remainder of this section, u_j^n denotes $u(j\Delta x, n\Delta t)$ and E_j^n denotes $U_j^n - u_j^n$ the error of approximation. We employ E^n to denote the vector of errors $(E_1^n, \ldots, E_J^n)^T$ and r^n to denote the vector of truncation error residuals $(r_1^n, \ldots, r_J^n)^T$ where

$$r_j^n = \frac{u_j^{n+1} - h(u_{j-1}^n, u_j^n)}{\Delta t}.$$

Furthermore, we let

$$R = \max_{0 \leq n\Delta t < \infty} \|r^n\|_\infty.$$

We first prove a preparatory lemma.

LEMMA 2.4. *Let $0 \leq j\Delta x \leq L$. Let $u_j^n, U_j^n \in [1-\epsilon, 1]$. Then, for ϵ sufficiently small,*

$$\|E^{n+1}\|_\infty \leq (1 - \bar{c}\Delta t)\|E^n\|_\infty + \Delta t\|r^n\|_\infty \tag{2.1}$$

for some constant $\bar{c} > 0$.

Proof. From the definition of truncation error and from (1.3) we find that

$$E_j^{n+1} = cE_{j-1}^n + (1-c)E_j^n + \Delta t[g(U_{j-1}^n, U_j^n) - g(u_{j-1}^n, u_j^n)] - \Delta t r_j^n.$$

Using the mean value theorem we have that, for some $t \in [0,1]$,

$$\begin{aligned} &g(U_{j-1}^n, U_j^n) - g(u_{j-1}^n, u_j^n) \\ &\quad = g_u(u_{j-1}^n + tE_{j-1}^n, u_j^n + tE_j^n)E_{j-1}^n + g_v(u_{j-1}^n + tE_{j-1}^n, u_j^n + tE_j^n)E_j^n. \end{aligned}$$

Applying the mean value theorem again shows that

$$\begin{aligned} &g(U_{j-1}^n, U_j^n) - g(u_{j-1}^n, u_j^n) \\ &\quad = E_{j-1}^n[g_u(1,1) + c_1(U_{j-1}^n - 1) + c_2(u_{j-1}^n - 1) \\ &\qquad + c_3(U_j^n - 1) + c_4(u_j^n - 1)] + E_j^n[g_v(1,1) + c_5(U_{j-1}^n - 1) + c_6(u_{j-1}^n - 1) \\ &\qquad + c_7(U_j^n - 1) + c_8(u_j^n - 1)] \end{aligned}$$

where the c_i are constants. Taking the infinity norm gives

$$\begin{aligned} \|E^{n+1}\|_\infty \leq |1 + \Delta t f'(1) + \Delta t[&c_1^*|1 - U_{j-1}^n| + c_2^*|1 - u_j^n| + c_3^*|1 - U_j^n| \\ &+ c_4^*|1 - u_j^n|]\|E^n\|_\infty + \Delta t\|r^n\|_\infty, \end{aligned}$$

where the c_i^* are constants. Using the fact that ϵ may be taken as small as required, the result follows from (III). □

THEOREM 2.5. *Let $0 \leq j\Delta x \leq L$. Then, assuming that $u(x,t) \in C^{2,2}([0,L] \times [0,\infty))$, we have for $\Delta t, \Delta x$ sufficiently small,*

$$\|E^{n+1}\|_\infty \leq \bar{c}[\Delta t + \Delta x]$$

for all $0 \leq n\Delta t < \infty$ *where* $\bar{c}$ *depends upon* L *but is independent of* $n\Delta t$.

Proof. Let T be the time such that the true solution satisfies $u(x,t) \in [1-\epsilon/2, 1]$ for $0 < x < L$ and $t \geq T$. Standard convergence arguments based on the maximum norm show that the conclusion of Theorem 2.5 is true for $0 \leq n\Delta t \leq T$. Let $N = [T/\Delta t]$. Then for $\Delta t, \Delta x$ sufficiently small, from this convergence result and Lemma 2.1, $U_j^n \in [1-\epsilon, 1]$ for $0 \leq j\Delta x \leq L$ and $N \leq n < M$, where M is the first integer greater than N such that $U_j^n < 1-\epsilon$ or $M = \infty$ if no such integer exists. We now show that $M = \infty$. Assume, for the purposes of contradiction, that M is finite. For $N \leq n < M$ the conditions of Lemma 2.4 hold. Iterating inequality (2.1) using the Gronwall lemma gives

$$\|E^n\|_\infty \leq e^{-\bar{c}\Delta t(n-N)}\|E^N\|_\infty + [1 + e^{-\bar{c}\Delta t(n-N)}]\frac{R}{\bar{c}}. \tag{2.2}$$

By choosing $\Delta t, \Delta x$ sufficiently small independently of $n\Delta t$ it is clear that $\|E^M\|$ can be made as small as required (using the regularity of the true solution to ensure that the truncation error R is $\mathcal{O}(\Delta t, \Delta x)$. Hence U_j^M remains in $[1-\epsilon, 1]$. This is the required contradiction and thus $M = \infty$. The final result now follows from (2.2), noting that n may be taken arbitrarily large. □

3. Asymptotic behaviour on an infinite spatial interval and travelling waves. In the previous section we proved a convergence result which was uniformly valid on arbitrarily large time intervals, provided that the spatial interval is fixed. However, typical solutions for the partial differential equation (PDE) involve wave fronts which propagate with a particular speed, and hence most of the structure of interest leaves any fixed spatial domain after a finite time. In this section we focus on the sense in which the numerical method captures this propagation. The difficulty with obtaining error estimates that are uniformly valid in space is that a tiny error in the approximation of the speed of propagation of the front can lead to an $\mathcal{O}(1)$ error since the true front and the numerical front are out-of-phase. As we now show, this can be overcome if the numerical solution is considered to be the solution of a slightly perturbed version of the original problem.

We start with a discussion of the true travelling waves in the PDE. We seek a solution of (1.1) with the form $u(x,t) = v(\xi)$ for $\xi = l(x - st)$. Here s is the wave-speed and l is a length-scale factor whose role is merely to simplify notation in the following material. Substituting this ansatz into (1.1) we obtain

$$\frac{dv}{d\xi} = \frac{f(v)}{l(1-s)}. \tag{3.1}$$

The properties of this scalar ordinary differential equation (ODE) are determined by (I)–(III). For each $s > 1$ the equation has a unique (up to translations) heteroclinic orbit connecting 1 at $\xi = -\infty$ to zero at $\xi = +\infty$. (This follows automatically for any $f(u)$ satisfying (I)–(III) since the equation is a scalar.) For $s < 1$ the connection is reversed. Thus waves exist for all values of the speed s. Which particular wave-speed is observed depends upon the decay of the initial data at infinity. If (3.1) is linearised about zero we find that

$$\frac{dv}{d\xi} \approx \frac{\mu}{l(1-s)},$$

where $\mu = f'(0)$. Hence, for speeds $s > 1$, the initial data must decay in x like $e^{\mu x/(1-s)}$.

The case $s = 1$ is a useful illustration of what is to follow. We show that, for a particular scheme with particular initial data, the scheme is uniformly convergent in both time and space, to a *perturbed* version of the original problem where the size of the perturbation is of $\mathcal{O}(\Delta t)$. The travelling wave is a discontinuity with the form

$$u(x,t) = 1, \qquad x < t,$$

$$u(x,t) = 0, \qquad x > t.$$

Now consider the numerical scheme (1.3) with the choice $g(u,v) = f(u)$. Thus the scheme becomes

$$U_j^{n+1} = cU_{j-1}^n + (1-c)U_j^n + \Delta t f(U_j^n).$$

Consider refining the mesh in such a way that $c + b\Delta t = 1$ for some fixed constant b. (By choosing b sufficiently large the monotonicity conditions (iii), (iv) will be satisfied.) We now seek a travelling wave solution of the form $U_j^n = V_k$ for $k = j - n$. We obtain

$$V_{k-1} = cV_{k-1} + (1-c)V_k + \Delta t f(V_k).$$

Using the fact that $1 - c = b\Delta t$, we obtain

$$V_{k-1} = V_k + \frac{f(V_k)}{b}. \tag{3.2}$$

Assume that (3.2) has a heteroclinic orbit satisfying $V = 1$ at $k = -\infty$, $V = \frac{1}{2}$ at $k = 0$, and $V = 0$ at $k = \infty$ and that this sequence is monotonic. (Such a connection exists for $f(u) = u - u^2$, for example.) Let $\hat{v}$ be the discontinuous solution of the differential equation centred at the origin so that

$$\hat{v} = 1, \qquad x < 0,$$

$$\hat{v} = 0, \qquad x > 0.$$

As the mesh is refined such that $c + b\Delta t = 1$ the heteroclinic orbit satisfying (3.2) converges to $\hat{v}$ in the topology of $L_1(\mathcal{R})$. The resultant wave moves one grid point in space in every time step. Hence its speed is $1/c = 1/(1 - b\Delta t)$. If we take the heteroclinic orbit satisfying (3.2) as initial data and refine the mesh, then it remains uniformly close in space and time to the discontinuous solution travelling with speed $1/(1 - b\Delta t)$. Thus the numerical solution corresponding to this *particular initial data and mesh-refinement path* is uniformly close in space and time to the solution of the equation

$$u_t + \frac{1}{1 - b\Delta t} u_x = f(u).$$

This is the original equation with an $\mathcal{O}(\Delta t)$ perturbation to the coefficients.

We now consider the general problem of the existence of discrete travelling waves and the relationship between their speed of propagation and that of the underlying continuous problem. The existence of numerical travelling waves for a scalar conservation law with no source term (such as Burgers' equation) was studied by [Jennings,

1974]. However, the method of analysis we employ is not the same since the stability properties of the points at $\pm\infty$ are entirely different. (It is possible that the method of Jennings might apply to the case $m < l$ which we do not consider here.)

We define an $m - l$ wave to be a wave which propagates m spatial grid points for every l temporal steps. Without loss of generality we assume that m and l are relatively prime. Specifically, we seek a wave in the form $U_j^n = V_k$ for $k = jl - mn$. Under this assumption (1.3) yields

$$V_{k-m} = cV_{k-l} + (1-c)V_k + \Delta t g(V_{k-l}, V_k). \tag{3.3}$$

Rearranging and setting $\Delta t = c\Delta x$ we obtain

$$\frac{V_{k-m} - cV_{k-l} - (1-c)V_k}{\Delta x(cl - m)} = \frac{cg(V_{k-l}, V_k)}{cl - m}.$$

Examination reveals that this forms a consistent approximation to the ODE (3.1) with the speed s given by $m/cl = m\Delta x/l\Delta t$. This agrees with our stipulation that the wave moves m grid points in l time steps. Note that the approximation may be in the form of a linear multistep method, a one-leg method or some hybrid of the two. It is natural to ask whether the numerical method also has heteroclinic orbits satisfying

$$V_{-\infty} = 1, \qquad V_{\infty} = 0. \tag{3.4}$$

Solutions of (3.3) satisfying this would represent numerical travelling waves. The following theorem shows that such waves do exist. In the subsequent analysis we refer to forward iteration as generating a sequence $\{V_k\}$ with $k \to \infty$ and backward iteration as generating a sequence $\{V_k\}$ with $k \to -\infty$. In the following statement, the one-parameter family corresponds to translation of the wave in space.

THEOREM 3.1. *Let $m > l$ be two relatively prime integers. For $0 < c < 1$ and Δt sufficiently small there exists a one-parameter family of solutions to* (3.3) *satisfying* (3.4).

The theorem is proved in a sequence of lemmas which we now present. The essence of the proof is as follows. The approximation (3.3) is zero-unstable for marching as $k \to \infty$ and zero-stable for $k \to -\infty$. We show that, for a carefully chosen set of initial points, the iteration (3.3) tends to 1 as $k \to -\infty$ and then that the carefully chosen set of initial points lie on the one-dimensional stable manifold of zero. The properties of the stable manifold follow from the zero-stability properties of the scheme.

LEMMA 3.2. *Let $m > l$ be two relatively prime integers. The polynomial*

$$\lambda^m - c\lambda^l - 1 + c = 0$$

has a root $\lambda = 1$ for all c and all other roots that lie inside the unit circle for $0 < c < 1$.

Proof. It is clear by inspection that $\lambda = 1$ is a root for all c. If $c = 0$ all roots lie on the unit circle and

$$\lambda = e^{2\pi ik/m}, \qquad k = 0, \ldots, m-1.$$

If $c = 1$ then $\lambda = 0$ is a root of multiplicity l and the remaining roots lie on the unit circle and satisfy

$$\lambda = e^{2\pi ik/(m-l)}, \qquad k = 0, \ldots, m-l-1.$$

By Theorem 6.41 of [Henrici, 1974] we deduce that, since $m > l$, 1 is an inclusion radius for the zeros of the polynomial for $0 < c < 1$ so that all roots satisfy $\lambda \leq 1$.

To complete the proof we now show that the roots cannot touch the unit circle for $0 < c < 1$. For the purposes of contradiction we assume that there is a root on the unit circle and we take

$$\lambda = e^{i\theta}$$

to obtain

$$e^{im\theta} - ce^{il\theta} = 1 - c.$$

Equating real and imaginary parts yields

$$\cos(m\theta) - c\cos(l\theta) = 1 - c,$$
$$\sin(m\theta) - c\sin(l\theta) = 0.$$

The first equation is equivalent to

$$\sin^2(m\theta/2) = c\sin^2(l\theta/2) \tag{3.5}$$

and the second implies

$$\sin^2(m\theta/2)\cos^2(m\theta/2) = c^2\sin^2(l\theta/2)\cos^2(l\theta/2). \tag{3.6}$$

If (3.5), (3.6) have no solution then neither do the original pair. Substituting (3.5) in (3.6) we obtain

$$c\sin^2(l\theta/2)[1 - c\sin^2(l\theta/2)] = c^2\sin^2(l\theta/2)[1 - \sin^2(l\theta/2)].$$

This implies that $c = 0$,

$$\sin^2(l\theta/2) = 0$$

or

$$1 - c\sin^2(l\theta/2) = c[1 - \sin^2(l\theta/2)].$$

Since we are studying $0 < c < 1$, the first option is not of interest and the third implies $c = 1$ which is also not of interest. The second implies $l\theta = 2k_1\pi$ for some integer k_1 which in turn implies $m\theta = 2k_2\pi$ because of (3.5). As k and l are relatively prime this implies that $k_2 = m$ and $k_1 = l$ so that $\theta = 2\pi$ and thus $\lambda = 1$ is the only possible root on the unit circle for $0 < c < 1$. Hence, since $\lambda = 1$ for all c, it remains to check that $\lambda = 1$ cannot be a multiple root.

A root of multiplicity $n > 1$ must satisfy the original polynomial plus the conditions

$$\begin{array}{ccccc} m & - & cl & = & 0, \\ m(m-1) & - & cl(l-1) & = & 0, \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m(m-1)\dots(m-n+2) & - & cl(l-1)\dots(l-n+2) & = & 0. \end{array}$$

Because $m > l$, $n \leq m$ and $0 < c < 1$ such conditions cannot be satisfied. This completes the proof. □

LEMMA 3.3. *Consider initial conditions for backward iteration of* (3.3) *satisfying* $0 < V_0 < V_{-1} < \ldots < V_{-m+1} < 1$. *Assume that these initial conditions generate* V_{-m} *satisfying* $1 > V_{-m} > V_{-m+1}$. *Then* $V_k \to 1$ *as* $k \to -\infty$.

Proof. We show, by induction, that the sequence generated is monotonic increasing as $k \to -\infty$. First note that $V_k, V_{k-1}, \ldots, V_{k-m+1} \in [0,1]$ implies that $V_{k-m} \in [0,1]$ by straightforward application of the monotonicity assumptions (iii) and (iv). By induction we deduce that $V_k \in [0,1]$ for $k \leq 0$. From (3.2) we have

$$V_{k-m} = h(V_{k-l}, V_k),$$
$$V_{k-m+1} = h(V_{k-l+1}, V_{k+1}).$$

Let $\delta_k = V_k - V_{k+1}$. Subtracting, and using the mean value theorem, we obtain

$$\delta_{k-m} = h_u(\xi, \eta)\delta_{k-l} + h_v(\xi, \eta)\delta_k.$$

From the assumption of the theorem we have $\delta_{-1}, \ldots, \delta_{-m} > 0$. By induction, noting that $\xi, \eta \in [0,1]$ and using (iii), (iv) we deduce that $\delta_k > 0$ for $k < 0$. Thus the sequence generated is monotonic increasing, remains in $[0,1]$, and has 1 as the only possible limit. Hence $V_k \to 1$ as $k \to -\infty$, and the proof is complete. □

In the following, it is convenient to consider (3.3) written as an implicit one-step map on $\mathcal{R}^m$. Let $\mathbf{W_k} = (V_{k-m+1}, \ldots, V_{k-1}, V_k)^T$, then (3.3) becomes

$$\mathbf{W_k} + \Delta t \mathbf{G}(\mathbf{W_k}) = B\mathbf{W_{k-1}}, \tag{3.7}$$

where

$$\mathbf{G}(\mathbf{W_k}) = (0, \ldots, \Delta t g(V_{k-1}, V_k)/(1-c))^T,$$

and B is an $m \times m$ matrix with zero entries everywhere except on the upper diagonal where it is 1, the bottom right-hand corner where it is $1/(1-c)$, and in the $(m-l+1)$th entry of the bottom row where it is $-c/(1-c)$. Equation (3.7) may have several solutions for $\mathbf{W_k}$. However, using the contraction mapping theorem, it is straightforward to show that (3.7) is uniquely solvable for $\mathbf{W_k}$ in a sufficiently small ball centred on the origin $\mathbf{O}$ in $\mathcal{R}^m$. Using this inverse we may write

$$\mathbf{W_k} = \mathbf{H}(\mathbf{W_{k-1}}) \tag{3.8}$$

for $\|\mathbf{W_k}\|$ sufficiently small. Equation (3.8) defines a C^1 diffeomorphism in a sufficiently small neighbourhood of the origin, with explicit inverse given by $\mathbf{W_{k-1}} = B^{-1}(\mathbf{W_k} + \Delta t \mathbf{G}(\mathbf{W_k}))$. The differentiability property follows from the smoothness of g.

LEMMA 3.4. *Consider the map* (3.8) *with* $m > l$ *and* $c \in (0,1)$. *Then, for* Δt *sufficiently small, the origin* $\mathbf{0}$ *has a one-dimensional stable manifold tangent to the eigenspace* $(\lambda^{m-1}, \lambda^{m-2}, \ldots, \lambda, 1)^T$, *for some* $\lambda > 1$.

Proof. We linearise the map (3.8) about $\mathbf{0}$ and look for characteristic multipliers of the linear equation by setting $\mathbf{W_k} = \lambda \mathbf{W_{k+1}}$. The linear problem involves a companion matrix and the characteristic polynomial is

$$\lambda^m - c\lambda^l - 1 + c - \Delta t g_u(0,0)\lambda^l - \Delta t g_v(0,0) = 0. \tag{3.9}$$

The linearised eigenspace corresponding to a particular root λ has the form $(\lambda^{m-1}, \ldots, 1)^T$. For $c \in (0,1)$ and Δt sufficiently small, we know from Lemma 3.2 that $m-1$ of

the roots lie inside the unit circle. By consistency, the root which is on the unit circle for $c \in (0,1)$ and $\Delta t = 0$ moves outside the unit circle, since zero is a stable steady solution of (3.1) for $s > 1$: seeking an expansion of this root for fixed c and $\Delta t \ll 1$ we find that

$$\lambda = 1 + \frac{\mu \Delta t}{m - cl} + \mathcal{O}(\Delta t^2). \tag{3.10}$$

Noting that $m > l$ and $c \in (0,1)$ it follows that $\lambda > 1$. Applying the stable manifold theorem for C^1 diffeomorphisms [Guckenheimer and Holmes, 1983] the result follows. □

Proof of Theorem 3.1. Consider the implicit map (3.3) written in the form (3.7). By Lemma 3.4 we can, by working sufficiently close to $\mathbf{0}$, choose a point $\mathbf{W_0}$ on the stable manifold of zero such that $0 < V_0 < V_{-1} < \cdots < V_{-m+1} < 1$. Furthermore, also by Lemma 3.4, the point $\mathbf{W_{-1}}$ generated by backward iteration of the map (3.7) will yield a point satisfying $1 > V_{-m} > V_{-m+1}$, again by ensuring that $\mathbf{W_0}$ is sufficiently close to $\mathbf{0}$. There is a one-parameter family of such initial conditions which may be found by varying $\mathbf{W_0}$ along the stable manifold from the origin until a point at which the above arguments do not hold.

With such a starting condition, Lemma 3.3 applies. Hence, under backward iteration of this starting value, the iterates of the map (3.7) will tend to $\mathbf{1}$ and we have found the one-parameter family of connecting orbits. □

We now derive a simple comparison principle showing that, under evolution, initial data trapped between two wave profiles remains trapped between them. This gives an estimate of the ultimate propagation speed for all initial data of this type. In the following it is useful to transform to a wave frame. We set $U_j^n = V_k^n$ for $k = jl - mn$ to obtain

$$V_{k-m}^{n+1} = cV_{k-l}^n + (1-c)V_k + \Delta t g(V_{k-l}^n, V_k^n). \tag{3.11}$$

THEOREM 3.5. *Consider the Cauchy problem* (3.11) *with initial data satisfying* $V_k^- \le V_k^0 \le V_k^+$ *where* V_k^-, V_k^+ *are two heteroclinic orbits satisfying* (3.3), (3.4). *Then* $V_k^- \le V_k^n \le V_k^+$ *for all* $n \ge 0$.

Proof. Let $\delta_k^n = V_k^n - V_k^-$. Subtracting (3.3) from (3.11) we obtain

$$\delta_k^{n+1} = h(V_{k-l}^n, V_k^n) - h(V_{k-l}^-, V_k^-).$$

By the mean value theorem we obtain

$$\delta_{k-m}^{n+1} = h_u(\xi,\eta)\delta_{k-l}^n + h_v(\xi,\eta)\delta_k^n.$$

From the proof of Theorem 3.1 it is known that the travelling wave $V_k^- \in [0,1]$ and, from a simple modification of Lemma 2.1 it may be shown that $V_k^n \in [0,1]$. Hence, by (iii), (iv) $h_u(\xi,\eta)$ and $h_v(\xi,\eta)$ are both positive. A straightforward induction then shows that the initial positivity of δ_k^n is preserved for all $n \ge 0$. A similar argument based on $V_k^+ - V_k^n$ completes the proof. □

For both the true and the numerical waves, the speed of propagation is governed by the decay of the initial data at infinity. Using this information, it is possible to study the relationship between the true and the numerical wave speeds. In the coordinate ξ, true waves with speed $s > 1$ decay like $\exp(\mu \Delta x / l(1-s))$ over an interval of length Δx at infinity. Numerical waves with speed m/lc decay like λ^{-1}, where λ is the unique root of (3.9) outside the unit circle (see Lemma 3.4). We now

assume that both problems are given initial data with the same decay rate at infinity and work out the relationship between the true speed s and the numerical speed m/lc. Matching the two decay rates we obtain

$$\lambda^{-1} = \exp(\mu \Delta x / l(1-s)).$$

Taking logarithms and rearranging gives

$$s = 1 + \frac{\mu \Delta t}{cl \log(\lambda)}.$$

We expand λ in powers of Δt for fixed c in $(0,1)$. Taking $\lambda = 1 + \lambda_1 \Delta t + \lambda_2 \Delta t^2 + \cdots$ we obtain

$$s = 1 + \frac{\mu}{cl\lambda_1} - \frac{\mu(\lambda_2 - \frac{1}{2}\lambda_1^2)}{cl\lambda_1^2}\Delta t + \cdots. \tag{3.12}$$

It may be shown that

$$\lambda_1 = \frac{\mu}{m - cl}$$

and that

$$\lambda_2 = \frac{1-(m+cl)}{2}\lambda_1^2 + \frac{l\mu g_u(0,0)}{(m-cl)^2}.$$

Hence

$$s = m/cl + \mathcal{O}(\Delta t). \tag{3.13}$$

This establishes that, for initial data with given decay rate, the numerical wave-speed m/cl converges to the true wave-speed s for initial data with given decay rate.

We now discuss what happens to the travelling waves as the stiffness of the source term is increased, for fixed values of c and Δt. We consider the case where $f(u)$ takes the form $f(u) = \mu h(u)$ for some source term normalised by $h'(0) = 1$. Increasing μ corresponds to increasing the stiffness of the problem. A geometric argument [Beyn, 1990] shows that the number of free parameters p expected to find a connecting orbit in m dimensions between a fixed point at $-\infty$ with unstable manifold of dimension u_- and stable manifold at $+\infty$ of dimension s_+ is

$$p = m + 1 - u_- - s_+.$$

In a typical situation, a single connecting orbit (up to translations) is to be expected if $p = 0$, a family of connections if $p < 0$, and no connections if $p > 0$. For example, the proof of Theorem 3.1 revolves around the fact that $u_- = m$ (which is why Lemma 3.3 holds) and $s_+ = 1$ (see Lemma 3.4) so that $p = 0$. Mere calculation of p does not constitute a proof, but nonetheless it is of interest to study how p varies with the stiffness of the problem, for fixed Δt. This gives an indication of what happens to the travelling waves constructed for Δt sufficiently small in Theorem 3.1.

A generalisation of the analysis leading to (3.9) shows that the roots of the polynomial

$$\lambda^m - [c + \Delta t g_u(\bar{u}, \bar{u})]\lambda^l - 1 + c - \Delta t g_v(\bar{u}, \bar{u}) = 0$$

govern the dimensions of the stable and unstable manifolds of a fixed point $\bar{u}$. The number of roots λ inside the unit circle gives the dimension of the unstable manifold and the number of roots outside gives the dimension of the stable manifold. We consider first the fixed point zero at $k = +\infty$. We let $g_u(0,0) = a\mu$ and $g_v(0,0) = b\mu$ where $a + b = 1$. As the problem becomes increasingly stiff $\mu \to \infty$. Thus, for highly stiff problems, the roots satisfy

$$\lambda^l \approx -b/a, \qquad \lambda^{m-l} \approx -\Delta t a \mu.$$

For $|b/a| > 1$ the dimension of the stable manifold at $k = +\infty$ is m and for $|b/a| < 1$ the dimension is $m - l$. A similar analysis at $k = -\infty$ where $\bar{u} = 1$ shows that for $|b/a| < 1$ the dimension of the unstable manifold is zero and for $|b/a| > 1$ it is l. Combining these facts shows that $p = 1$ whatever values b and a take so that a connection is not to be expected. On this basis we make the following conjecture.

CONJECTURE 3.6. *The travelling waves constructed in Theorem* 3.1 *cease to exist as the stiffness of the source term is increased, for fixed values of* $c, \Delta t$.

It is clear that conditions (iii) and (iv) on the scheme will be violated with increasing stiffness, for fixed $c, \Delta t$. Numerical experiments in §5 bear out Conjecture 3.6 and show how it is related to the loss of monotonicity caused by the violation of (iii), (iv).

4. The split-step scheme. In this section we describe and analyse a split-step scheme, mainly for the purposes of comparison with the scheme (1.3). The scheme is based on exact integration along the characteristics, with initial values between grid points determined by interpolation. Consider the equation

$$u_t = f(u), \qquad u(0) = u_0.$$

We introduce the evolution semigroup $S(t)$ to denote the solution $u(t) = S(t) \bullet u_0$. Employing the same notation as in §1, the split-step scheme comprises the two steps

$$\begin{aligned} U_j^* &= cU_{j-1}^n + (1-c)U_j^n, \\ U_j^{n+1} &= S(\Delta t) \bullet U_j^*. \end{aligned} \tag{4.1}$$

The only requirement that we make on the scheme is that $c \in [0,1]$. It is then automatic that $U_j^0 \in [0,1]\ \forall_j$ implies $U_j^n \in [0,1]\ \forall_j$ and $\forall n \geq 0$.

As in §3 we seek $m - l$ travelling waves by setting $U_j^n = V_k$ for $k = jl - mn$. This gives the defining equation

$$V_{k-m} = S(\Delta t) \bullet (cV_{k-l} + (1-c)V_k) \tag{4.2}$$

together with condition (3.4) for a heteroclinic connection. The following result may be proved similarly to Theorem 3.1, and we only sketch the proof.

THEOREM 4.1. *Let* $m > l$ *be two relatively prime integers. For* $0 < c < 1$ *and* Δt *sufficiently small there exists a one-parameter family of solutions to* (4.2) *satisfying* (3.4).

Sketch proof. Consider (4.2) as an explicit map in the form (3.8). This may be done by inverting $S(\Delta t)$ in (4.2) to obtain $(1-c)V_k = S(-\Delta t) \bullet V_{k-m} - cV_{k-l}$ and formulating this as a one-step map on $\mathcal{R}^m$. It is possible to show that Lemma 3.4 holds for (4.2) rather than (3.3): the characteristic equation replacing (3.9) is

$$\lambda^m - \exp(\mu \Delta t)(c\lambda^l + 1 - c) = 0. \tag{4.3}$$

and straightforward calculation shows that (3.10) again holds. Thus the stable manifold of **0** is one-dimensional and points on it satisfy $0 < V_0 < V_{-1} < \cdots < V_{-m+1} < V_{-m} < 1$. The map (4.2) may now be iterated backwards, starting on this stable manifold. Lemma 3.3 may be generalised from (3.3) to (4.2) since $S(\Delta t) \bullet u$ is monotone: it satisfies

$$\frac{d}{du}(S(\Delta t) \bullet u) = \exp\left(\int_0^{\Delta t} f'(S(\tau)u)\right) d\tau.$$

Thus $V_k \to 1$ as $k \to \infty$, and the proof is complete. □

Analysis of (4.3) reveals that, for the split-step scheme, the true and numerical wave-speeds are related by the expression (3.12) where now

$$\lambda_2 = \frac{1-(m+cl)}{2}\lambda_1^2 + \frac{cl\mu^2}{(m-cl)^2} + \frac{\mu^2}{2}.$$

As for the explicit scheme in §3, we can examine the existence of the travelling waves as the stiffness of the problem increases, for fixed $c, \Delta t$. Similar arguments may be made, based on the polynomial

$$\lambda^m - \exp(f'(\bar{u})\Delta t)(c\lambda^l + 1 - c) = 0.$$

These arguments show that, for $m > l$, $p \le 0$ for μ large with $c, \Delta t$ fixed. Hence connections remain possible as the stiffness increases without bound. Since Lemma 3.3 holds independent of stiffness, the existence of connections for all values of stiffness could be made rigorous simply by establishing that (4.3) has at least one positive real root for all μ, however large. Extensive numerical computations reveals this to be true. Thus the situation is very different from the explicit scheme considered in §3. This is borne out in the numerical experiments described in the next section.

5. Numerical results. In this section we describe some numerical schemes satisfying the conditions (i)–(iv) and present some numerical results. The particular scheme we choose involves the function

$$g(u,v) = \theta f(\gamma u + (1-\gamma)v) + \phi f(u) + (1-\theta-\phi)f(v). \tag{5.1}$$

Here the parameters satisfy $\theta, \gamma, \phi, \theta+\phi \in [0,1]$. Let

$$K_1 = -\min_{u\in[0,1]} f'(u).$$

Then the monotonicity conditions (iii), (iv) become, for $\alpha = \theta\gamma + \phi$,

$$\Delta t \alpha K_1 \le c,$$

and

$$c + \Delta t[1-\alpha]K_1 \le 1. \tag{5.2}$$

The choice $\alpha = 1 - c$ corresponds to picking up information from along the characteristic. Note that stiff problems, for which $K_1 \gg 1$, typically involve severe restrictions on both Δt and Δx.

Throughout this section we take $f(u) = \mu u(1-u)$. As in [LeVeque and Yee, 1990] we define the numerical wave-speed by

$$speed = \frac{1}{c}\left(\sum_j U_j^n - \sum_j U_j^{n-1}\right).$$

This definition is appropriate given the expected form of the solution. All our numerical computations concerning estimates of the wave speed were checked independently using an alternative definition based on locating the level set $u(x,t) = \frac{1}{2}$ in space-time by interpolation between the grid points. In all the figures the initial data is an exact travelling wave solution of (C) corresponding to wave-speed 2.

Figures 1–3 concern the scheme (1.3), (5.1) with $\theta = \phi = \gamma = 0$. Figure 1 shows numerically computed solution profiles. The results clearly show the devolopment of a wave-profile with constant speed of propagation. This reflects the existence Theorem 3.1. Figure 2 shows numerical estimates of the wave-speed calculated in four computations along a mesh-refinement path satisfying $c = 0.5$. The linear convergence of the wave-speed to its true value of 2 is clear. This reflects the relationship (3.13) between the true and numerical wave-speeds. Figure 3 shows what happens when the monotonicity conditions are violated: for the computation shown $\alpha = 0$ and $c + \Delta t K_1 = 1.75$. Thus (5.2) is violated. The profiles are graphed at intervals of four seconds, starting at $t = 4$. The solution develops oscillations around the front and trails a structure reminiscent of the bifurcation diagram for the quadratic map: the solution profile has (approximately) period 2 in space, then period 4, etc.

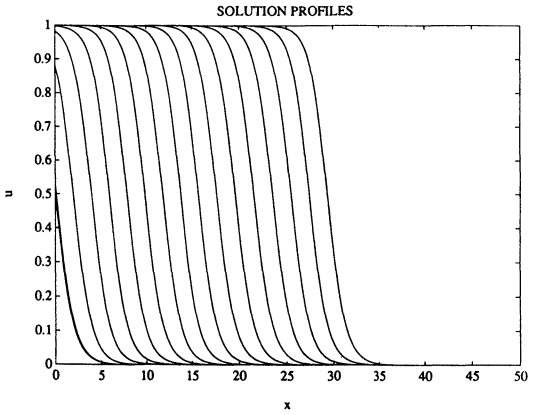

FIG. 1. *Scheme* (1.3). $\mu = 1, c = 0.5, \Delta t = 0.05$. *Profiles every second.*

Figures 4–6 concern the split-step scheme (4.1). Figures 4 and 5 are analogous to Figs. 1 and 2 and they illustrate similar phenomena. Figure 6 illustrates the occurence of spurious wave-speeds in highly stiff problems [Colella et al., 1986; LeVeque and Yee, 1990]: after some time the solution is attracted to the profile which propagates with speed $1/c$: that is, one grid point per time step.

Acknowledgments. This work was started whilst D.F. Griffiths and A.M. Stuart were visiting scientists at NASA Ames Research Center. The authors are grateful to Peter Sweby for useful advice during this visit.

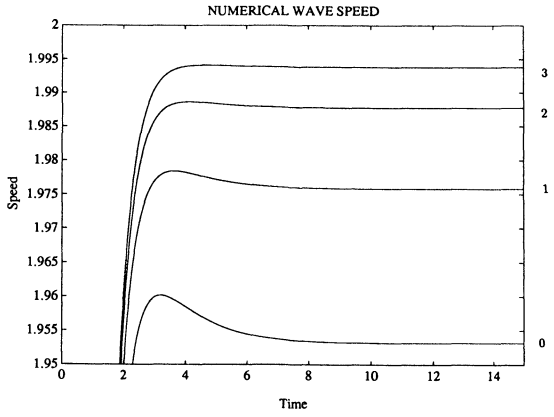

FIG. 2. *Scheme* (1.3). $\mu = 1, c = 0.5, \Delta t = 0.05/2^p, p = 0, 1, 2, 3.$

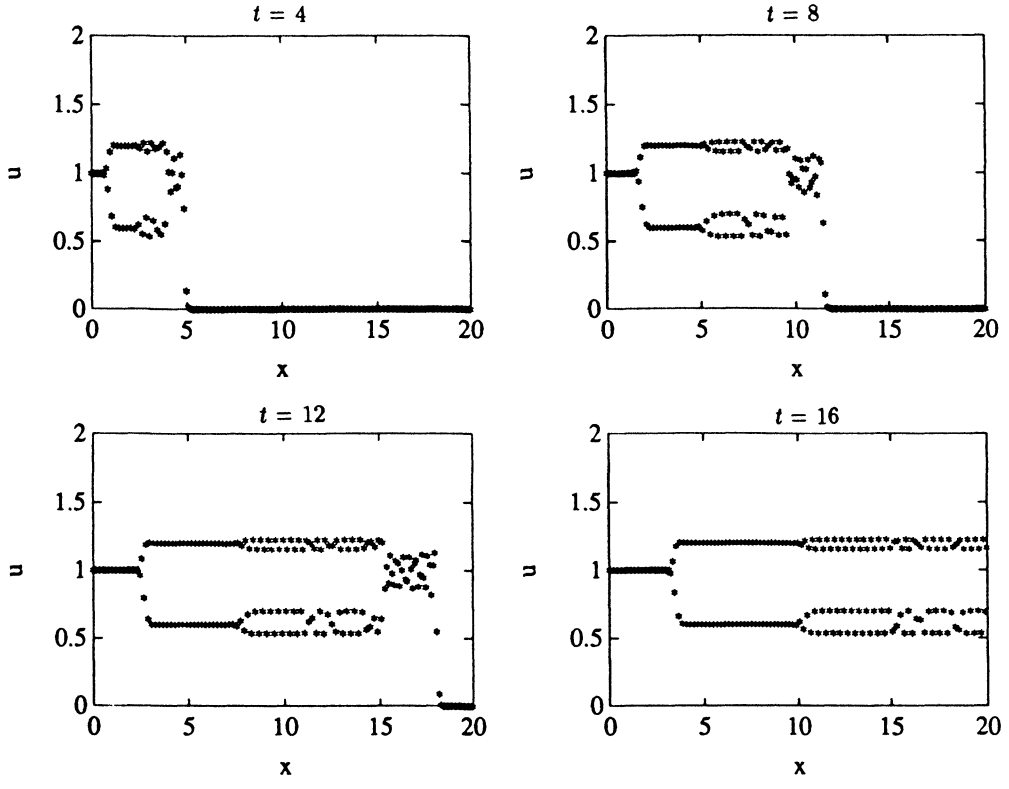

FIG. 3. *Scheme* (1.3). $\mu = 25.0, c = 0.5, \Delta t = 0.05$. *Profiles every four seconds.*

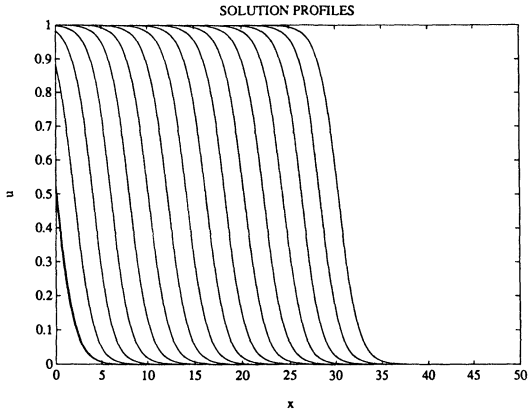

FIG. 4. *Scheme* (4.1). $\mu = 1, c = 0.5, \Delta t = 0.05$. *Profiles every second.*

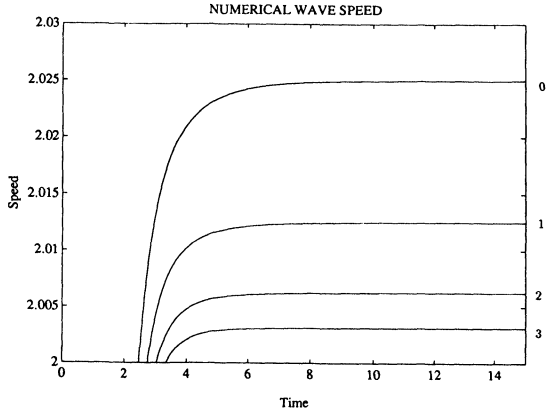

FIG. 5. *Scheme* (4.1). $\mu = 1, c = 0.5, \Delta t = 0.05/2^p, p = 0, 1, 2, 3$.

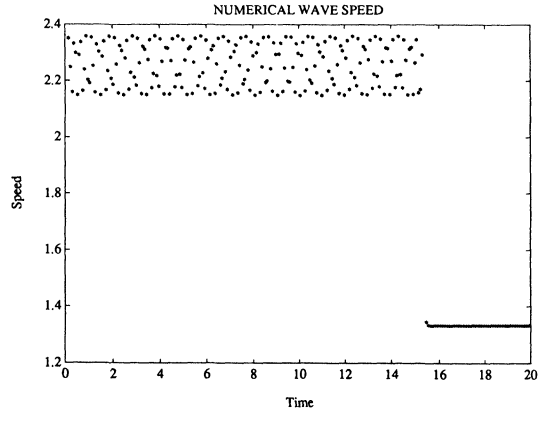

FIG. 6. *Scheme* (4.1). $\mu = 50.0, c = 0.75, \Delta t = 0.075$.

REFERENCES

[1] W. J. BEYN, *The numerical computation of connecting orbits in dynamical systems*, IMA J. Numer. Anal., 10 (1990), pp. 379–406.

[2] P. COLELLA, A. MAJDA, AND V. ROYTBURD, *Theoretical and numerical structure for reacting shock waves*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1059–1080.

[3] J. GUCKENHEIMER AND P. HOLMES, *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*, Appl. Math. Sci. 42, Springer-Verlag, New York, 1983.

[4] A. HARTEN, J. M. HYMAN, AND P. D. LAX, *On finite-difference approximations and entropy conditions for shocks*. Comm. Pure and Appl. Math., 29 (1976), pp. 297–232.

[5] P. HENRICI, *Applied and Computational Complex Analysis—Vol.* 1, Wiley Interscience, 1974.

[6] G. JENNINGS, *Discrete shocks*, Comm. Pure and Appl. Math., 27 (1974), pp. 25–37.

[7] S. LARSSON, *The long-time behaviour of finite-element approximations of solutions to semilinear parabolic problems*, SIAM J. Numer. Anal., 26 (1989), pp. 348–365.

[8] R. LEVEQUE AND H. C. YEE, *A study of numerical methods for hyperbolic conservation laws with source terms*. J. Comp. Phys., 86 (1990), pp. 187–210.

[9] J. M. SANZ-SERNA AND A. M. STUART, *A note on uniform in time error estimates for approximations to reaction-diffusion equations*, IMA J. Numer. Anal., to appear.